Post-Processing Data QA/QC

by

Taiye B. Sangoyomi, Ph.D., P.E.

Presentation Outline

- **What is post-processing QA/QC?**

- **Why do it?**

- **Where is it done?**

- **How is it done?**

What is Post-Processing QA/QC?

- Rigorous examination of data to ascertain and/or improve its quality

- Done after the data processing/initial QA-QC and flow computations

- Global QA/QC

- Checks the entire data production process, from data measurement, collection, transmission, processing and flow computations/streamgaging

Why do Post-Processing QA/QC?

- Millions of data records are collected and posted to DBHydro after data processing, initial QA/QC and flow computation

- Post-processing QA/QC performed on a subset of those data to meet various legally mandated data requirements, such as

  * Everglades Agricultural Area Rulemaking.
  * Stormwater Treatment Areas
  * Chapter 40E-63, Florida Administrative Code
  * Everglades Construction Project

- A large percentage of QA/QC functions are committed to fulfilling these data needs

- Lack of resources prevent post-processing QA/QC of entire District data set

- Currently making proposal to do this

Why? (contd)

Used to support many District activities, including:

- **Water supply, water budget, and water quality analyses**
- **Flood plain studies, flood control planning, and flood frequency analyses**
- **Hydrologic modeling**
- **Assessment of ecological restoration efforts, and**
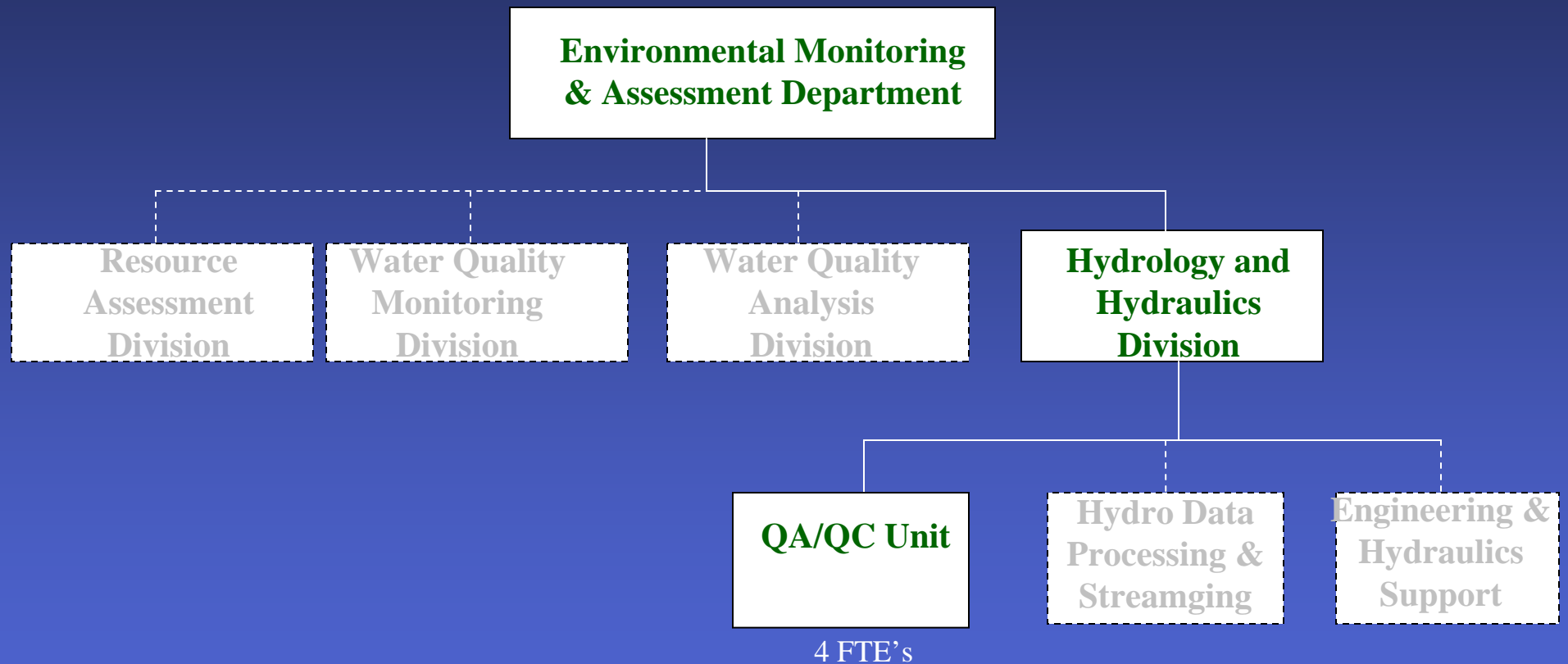- **Design of new water control structures**

These needs are met by providing what is known as "preferred data"

Preferred Data
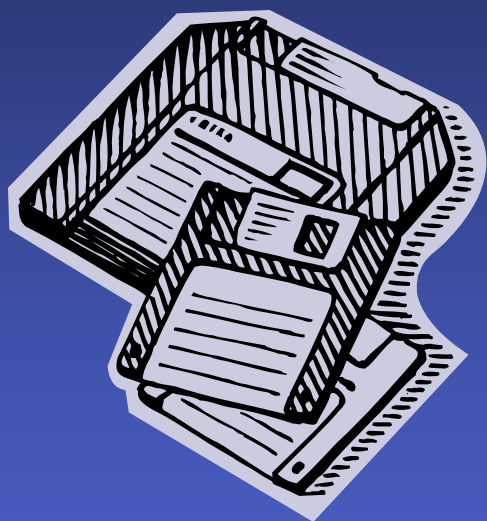
- Best available data

- Composed of the most appropriate combination of data available from any known data source

- Production is accomplished through a series of data and statistical analyses.

SOUTH FLORIDA WATER MANAGEMENT DISTRICT

Where is it done?

Environmental Monitoring & Assessment Department

- Resource Assessment Division
- Water Quality Monitoring Division
- Water Quality Analysis Division
- **Hydrology and Hydraulics Division**
  - **QA/QC Unit**

    4 FTE's
  - Hydro Data Processing & Streamging
  - Engineering & Hydraulics Support

sfwmd.gov

How?

Data Detective

☐ **Investigate data records to detect anomalies**

☐ **Correct them and/or prevent them from occurring again**

☐ **Forensic Hydrology**

How?

Tools

- Site knowledge
- Application of engineering know how
- Comparison to historical patterns
- Graphical inspection
- Communication

sfwmd.gov

How?

Tools (contd)

- Statistical analysis
- Alternate data sets
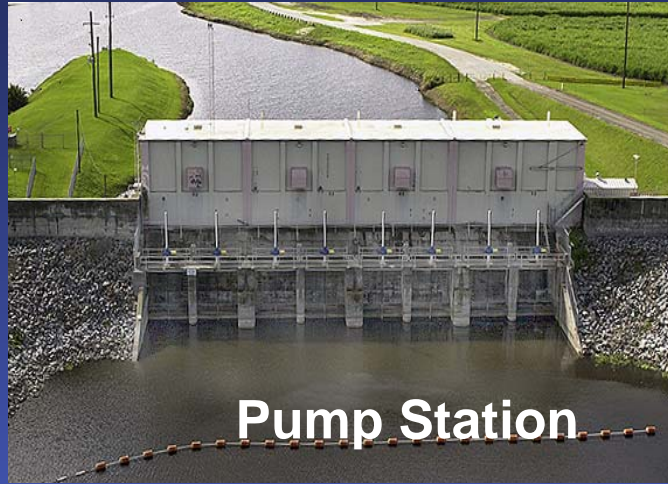- Mass balance analysis

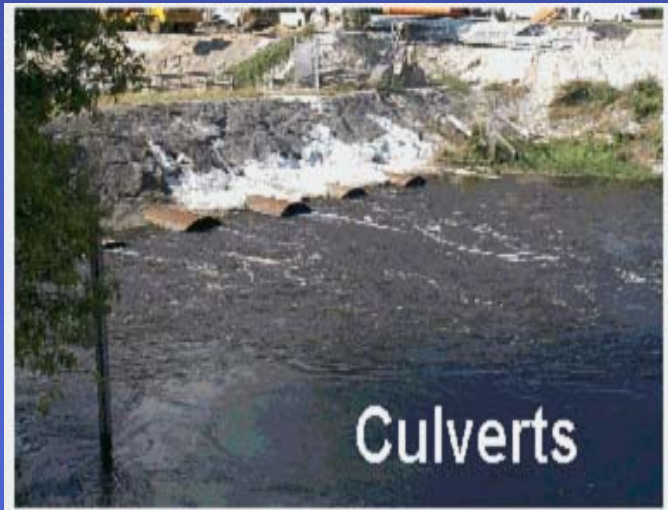# Knowledge of Site

Gated Spillway

Pump Station

Rain-gage

Culverts

Weirs
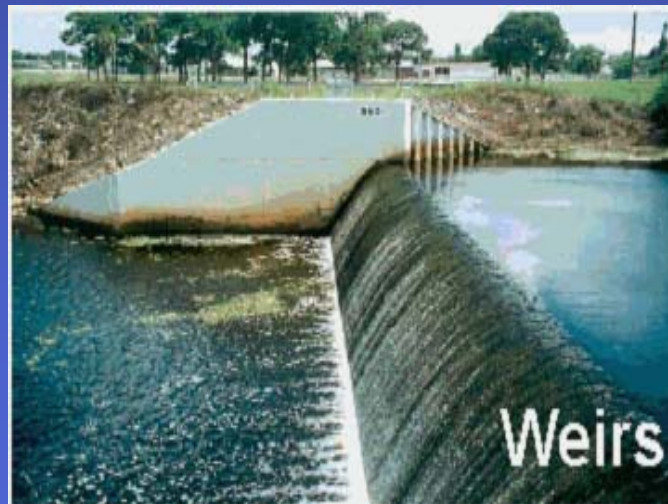
Stilling Well

Application of Engineering Know How

- **Does data make sense?**

- **Negative flow, is it possible at site?**

- **Is particular stage value realistic?**

- **Use hydraulic/hydrology engineering know how**

Historical Patterns

S-18C USGS and SFWMD FLOW
(1985- 2003)
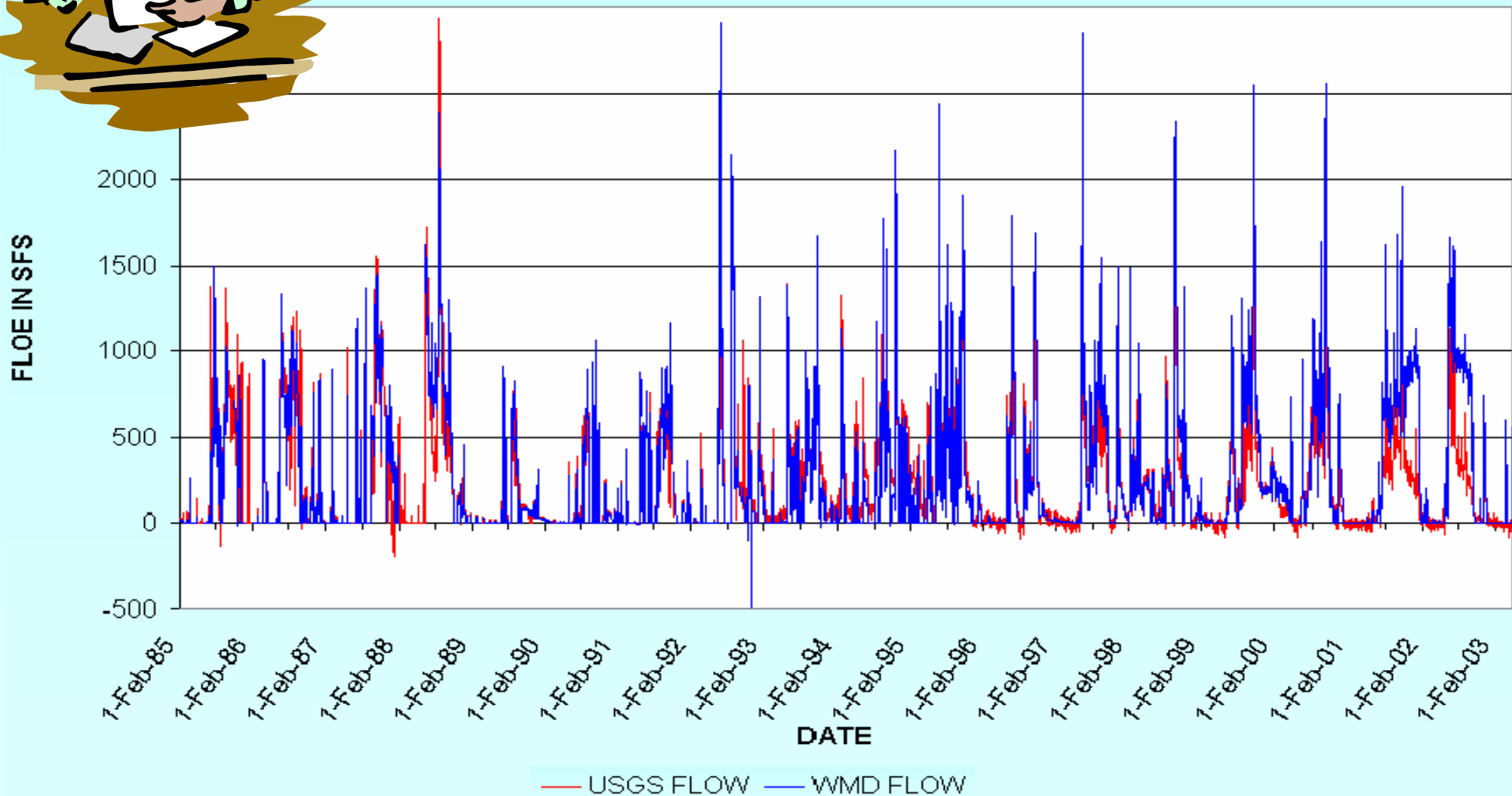
Graphical Analysis

S-18C (USGS vs. SFWMD) and (S-177 + S-178 total) FLOW COMPARISION
With
(Stage and Rainfall)

Legend:
- S-18C (STAGE) H.W
- S-18C (STAGE) T.W
- S-177 (STAGE) T.W
- S-18C_ RAINFALL
- S - 18C (USGS) Flow
- S-18C (WMD) FLOW
- S-177+ S-178 TOTAL FLOW
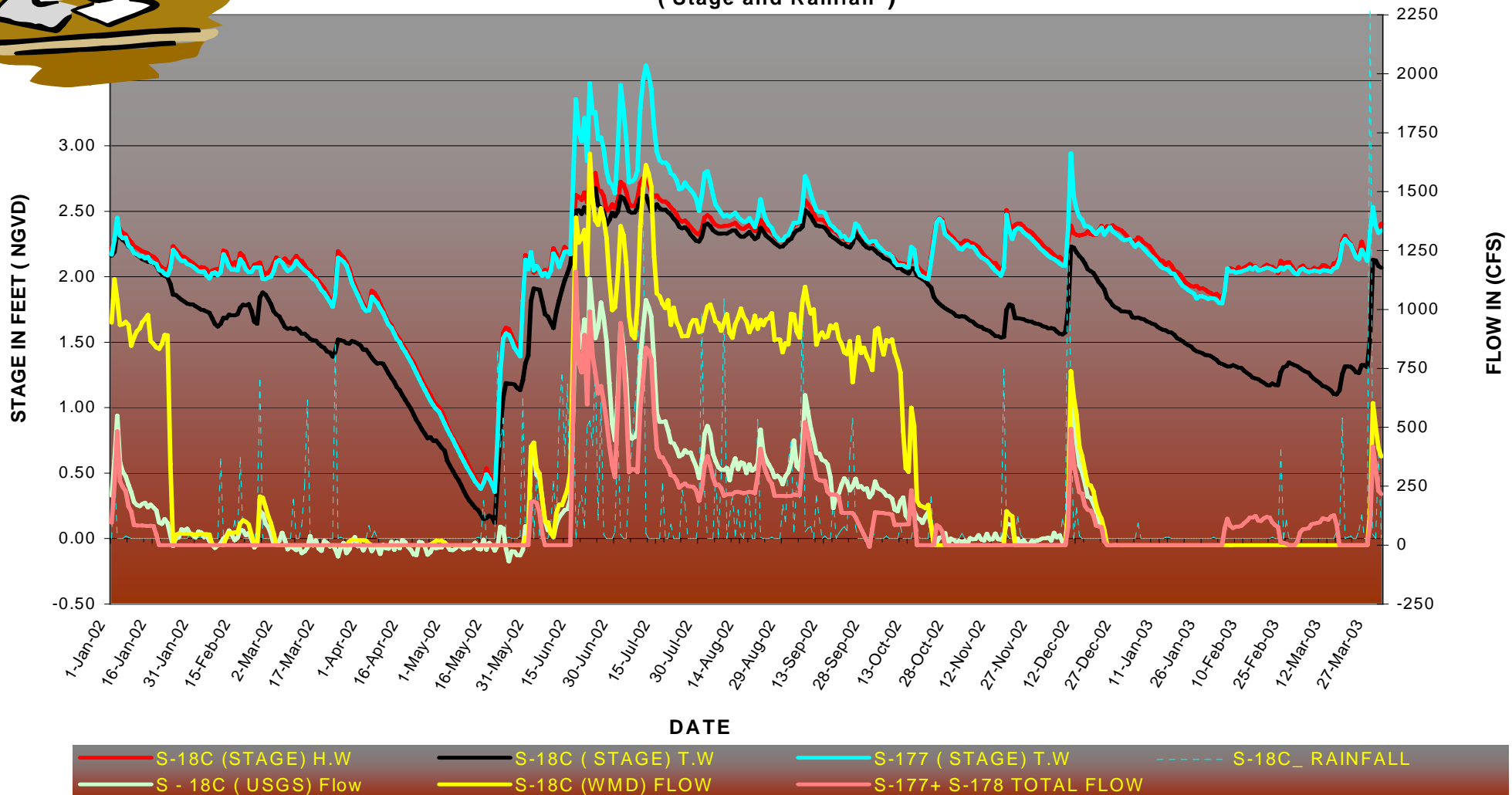
Communication

To find out what went wrong, how it went wrong & what can be done to correct it

- **Communicate with data processing staff**

- **Communicate with field personnel**

- **Communicate with hydraulic engineers**

Statistical Analysis

- **Regression**
- **Correlation**
- **Data Estimation Techniques**
  - **Regression, interpolation, direct substitution, inverse distance relationships**

Alternate Data Sources

For comparison and substitution

- **District data**

- **USGS data**

- **COE data**

- **NOAA data**

- **Other agency data**

Mass Balance Analysis

- Inflow = Outflow
             + Change in Storage

Flow Chart of QA/QC Procedures for Flow Data

Procedure Inputs

Procedure Steps

Procedure Outputs

Procedure Inputs

- USGS, COE, NOAA, Other Agency data
- DBHydro source data
- SFWMD processed data
- DCVP

Procedure Steps

1. Familiarize with structure, maximum flows and discharge characteristics
2. Review historical data, identify historical flow patterns
3. Validate Data. Graph data, identify missing values and outliers. Detect anomalies by comparing data to historical patterns, through mass balance analysis, and using statistical techniques such as correlations and regressions with nearby stations
4. Find out reason(s) for anomalies, missing and outlier data by looking at MIRMAID reports, annotations, and communicating with field, data processing, and engineering support staff
5. Fix problem at source of bad data from review carried out, if possible, e.g., from incorrect instrument calibration
6. Perform flow calculations using flow equations. If possible, derive flow values from independent equations to verify flow values in DBHydro
7. Estimate bad, missing, or outlier values using regression and other statistical methods, as well as using alternate data sources
8. Create continuous quality data set for station with estimated and alternate data sources. Place E (estimated) tags beside estimated data
9. Move quality assured data series to Preffered DBKeys
10. Load data into DBHydro
11. Document the steps, analysis, and results from QA/QC

Procedure Outputs

- DBHydro with preferred data set
- QA/QC analysis documentation

Shape Legend

- Stored Data
- Database
- Process
- Document

Questions